

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## *D*-optimal designs for multiarm trials with dropouts

### Journal Item

#### How to cite:

Lee, Kim May; Biedermann, Stefanie and Mitra, Robin (2019). *D*-optimal designs for multiarm trials with dropouts. *Statistics in Medicine*, 38(15) pp. 2749–2766.

For guidance on citations see [FAQs](#).

© 2019 The Authors



<https://creativecommons.org/licenses/by/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:  
<http://dx.doi.org/doi:10.1002/sim.8148>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

RESEARCH ARTICLE

# D-optimal designs for multiarm trials with dropouts

Kim May Lee<sup>1</sup>  | Stefanie Biedermann<sup>2</sup> | Robin Mitra<sup>3</sup>

<sup>1</sup>MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

<sup>2</sup>School of Mathematical Science, University of Southampton, Southampton, UK

<sup>3</sup>Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

## Correspondence

Kim May Lee, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SR, UK.  
Email: kim.lee@mrc-bsu.cam.ac.uk

## Present Address

Kim May Lee, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SR, UK.

## Funding information

Institute for Life Sciences at the University of Southampton; Medical Research Council, Grant/Award Number: MR/N028171 and MC\_UP\_1302/4; MRC and Alzheimer's Society UK, Grant/Award Number: RCTN49545035

Multiarm trials with follow-up on participants are commonly implemented to assess treatment effects on a population over the course of the studies. Dropout is an unavoidable issue especially when the duration of the multiarm study is long. Its impact is often ignored at the design stage, which may lead to less accurate statistical conclusions. We develop an optimal design framework for trials with repeated measurements, which takes potential dropouts into account, and we provide designs for linear mixed models where the presence of dropouts is noninformative and dependent on design variables. Our framework is illustrated through redesigning a clinical trial on Alzheimer's disease, whereby the benefits of our designs compared with standard designs are demonstrated through simulations.

## KEYWORDS

available case analysis, design of experiments, linear mixed models, noninformative dropouts

## 1 | INTRODUCTION

Clinical trials often involve several follow-up events for studying the impact of treatment regimes on subjects over the course of the trial. In recent years, multiarm designs have become more prominent than two-arm trials as the former design recruits one control group and several treatment arms in a single trial, leading to gains in efficiency for the overall study. However, the issues caused by the presence of dropouts are unavoidable in multiarm trials with follow-up, especially when the course of the study is long. It is common practice to tackle the presence of dropouts by scaling up the effective total sample size of the study, independently of the design. If the control group has significantly more dropouts than other treatment groups, statistical inference on pairwise treatment-control comparisons might still be distorted.

Assuming that repeated measurements on the same patients are independent, which may not be realistic, Galbraith et al<sup>1</sup> investigate power calculations for a longitudinal study in the presence of dropouts. In the literature on design of experiments, some authors<sup>2-5</sup> investigate the robustness of designs to missing values; others account for the presence of missing responses in the respective design criteria.<sup>6,7</sup> There are optimal design frameworks for regression

.....  
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

models in the presence of responses missing at random<sup>8,9</sup> and in the presence of responses not missing at random.<sup>10</sup> The attention on a design framework for multiarm studies with repeated measurements is rather limited, with only one study taking dropouts into account. However, this is a longitudinal study with only one group of participants to be followed-up throughout the course of the study.<sup>5</sup> Assuming completely observed data, several authors<sup>11–13</sup> focus on optimal design methodology for linear mixed models with a fixed number of time points for a longitudinal study.

A key feature of an optimal experimental design is its cost efficiency. For a fixed trial budget, an optimal design will provide the largest possible amount of information from the data. If a *D*-optimal design assuming all responses will be observed is used, it has been noted that “... researchers can easily compensate for such a small efficiency loss by increasing the sample size by at most 15%.”<sup>5</sup> However, in many situations a 15% increase in sample size can already incur considerable extra costs, and there are further ways to make a clinical trial design even more efficient. In our investigation, we redesign a study<sup>14</sup> assuming a fixed budget and realistic relative costs of recruiting a new participant versus measuring an existing participant at a further time point, thus finding the cost-optimal number of time points for the study in addition to the optimal locations of these time points.

We propose an optimal design framework for trials/studies that involve multiple groups of participants and several follow-up sessions, whereby linear mixed models with some pragmatic design constraints are considered as well as the presence of dropouts. By definition, a dropout refers to an experimental unit whose information is not being observed further once the outcome variable on the subject is not being measured at a time point. We consider noninformative dropout that depends on design variables, such as time of follow-up visits, in our investigation. We assume at the design stage of the study that a linear mixed model will be fitted to the incomplete data using the available subject data, ie, all observed responses of subjects during the study. This missing data analysis approach is appealing for its simplicity of application and as it makes inferences based on all the observed data. Moreover, likelihood-based inference yields valid conclusions in this framework when missing responses are noninformative.<sup>15–18</sup>

Our framework presents a more general approach to finding optimal designs over previously proposed approaches and allows differential missingness between different treatment groups. This framework also considers more complex design problems where a design consists not only of the optimal time points to measure subjects but also of other decision variables such as the number of time points (if using a fixed cost constraint) and the optimal allocation of dose levels of a drug. We have thus unified the approaches that consider a single cohort longitudinal study with dropouts,<sup>5</sup> without dropouts,<sup>11</sup> and studies with more than one group of participants,<sup>13</sup> while additionally incorporating cost and drug dose level considerations. In other words, our framework could provide optimal designs for longitudinal observational studies as well as for multiarm studies with follow-ups. Furthermore, our framework could include more sophisticated and pragmatic design considerations. In an unblinded trial setting, we investigate the designs where different groups of participants could have different time points versus designs where everyone is measured at the same time points. This work fills an important gap by incorporating practical design settings in the framework such that practitioners could find and explore the operating characteristics of various potential optimal designs before the implementation of a design. In addition, to our knowledge, there has been no optimal design framework that seeks to simultaneously optimize dose levels of a drug together with optimizing follow-up time point measurements. This is a key contribution of our work and would be useful for practitioners that seek to both learn about dose-response relationships and performance of different dose groups over time within a given study.

The structure of this paper is as follows. Section 2 introduces the concept of linear mixed models and the notion of dropout, as well as depicting the optimal design framework for multiarm studies that have more than one measurement observed on the same experimental units over the course of the trial. Section 3 then derives the optimal design framework in the presence of dropouts. Section 4 revisits a trial on Alzheimer's disease<sup>14</sup> to construct optimal designs through the derived framework, using the information from the study to elicit the dropout probability functions and the model parameters. Here, the main focus is on finding the design that provides the most information on a fixed budget. We use simulations to compare the performance of the designs. Section 5 concludes this paper with some discussion and research directions for future work.

## 2 | BACKGROUND AND NOTATIONS

### 2.1 | General linear mixed models

We now present the general formulation of a linear mixed model. Please note, while we define all notations used in the text at the time of introduction, for convenience, we also present a glossary in the Appendix that explains all the notations

used within this paper. Let  $\mathbf{y}_i^T = (y_{i1}, y_{i2}, \dots, y_{iq})$  be the  $q$  repeated measurements of subject  $i, i = 1, \dots, N$ . The responses of subject  $i$  can be represented by the linear mixed model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i,$$

where  $\boldsymbol{\beta}$  is a vector of unknown fixed parameters,  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  are design matrices,  $\mathbf{b}_i$  is a vector of unknown random coefficients that is normally distributed with mean zero and covariance matrix  $\mathbf{D}$ , ie,  $\mathbf{b}_i \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{D})$ , and  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{iq})^T$  is a vector of observational errors that is normally distributed with mean zero and covariance matrix  $\sigma^2 \boldsymbol{\Psi}$ , ie,  $\boldsymbol{\epsilon}_i \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \sigma^2 \boldsymbol{\Psi})$ . Moreover,  $\mathbf{b}_i$  and  $\boldsymbol{\epsilon}_i$  are assumed to be independent,  $i = 1, \dots, N$ . When we find an optimal design, we conjecture the structure of  $\boldsymbol{\Psi}$ , leading to a locally optimal design; see Section 2.4.

The maximum likelihood method provides an unbiased estimator for the fixed effect parameters  $\boldsymbol{\beta}$  of the linear mixed model, ie,  $\hat{\boldsymbol{\beta}} = (\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i$ , with covariance matrix  $\text{cov}(\hat{\boldsymbol{\beta}}) = (\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1}$ , where  $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \sigma^2 \boldsymbol{\Psi}$  is the covariance matrix of  $\mathbf{y}_i$  and  $\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i$  is called the Fisher information matrix. Note that this information matrix is summing the individual information that is being contributed by each experimental subject, and can be reexpressed as

$$\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i = \sum_{k=1}^c n_k \mathbf{X}_k^T \mathbf{V}_k^{-1} \mathbf{X}_k = N \sum_{k=1}^c w_k \mathbf{X}_k^T \mathbf{V}_k^{-1} \mathbf{X}_k, \quad (1)$$

where we assume that units are allocated into  $c$  groups with the same values of the design variables in each group. We can then denote  $X_k$  to be the unique design matrix of the  $k$ th group, where  $k = 1, \dots, c$ , with  $n_k$  ( $w_k = n_k/N$ ) reflecting the number (proportion) of experimental units in group  $k$ , respectively.

## 2.2 | Dropout mechanisms

We now illustrate the classification of dropout processes.<sup>19</sup> Let  $y_{ij}$  denote the  $j$ th measurement taken for subject  $i$ , where  $i = 1, \dots, N$  and  $j = 1, \dots, q$ . Similarly, denote a binary missing data indicator  $l_{ij}$  where  $l_{ij} = 1$  denotes  $y_{ij}$  is missing and  $l_{ij} = 0$  denotes  $y_{ij}$  is observed. The missing responses are said to be completely random drop-out if  $P(l_{ij} = 1)$  is a constant  $\forall i, j$ ; random drop-out if  $P(l_{ij} = 1)$  only depends on observed information; and informative drop-out if the drop-out process depends on unobserved measurements or the missing value itself. The presence of informative drop-out is more complicated than the presence of other processes and it often requires special treatment and sensitivity analysis based on the incomplete data. Techniques for the analysis of incomplete data with dropout are available.<sup>20</sup> We note that the analysis techniques are not the key interest of our investigation as we focus on finding a design prior to observing a study or an experiment. Nevertheless, the power of a study would be reduced if the dropout proportion is not accounted for at the design stage of the study.

In this paper, we investigate the role of noninformative, covariate, and group dependent dropout processes at the design stage of a study. Let  $n_{k,j}$  be the number of subjects in group  $k$  who remain in the experiment at time point  $j$ . A dropout process would cause  $n_{k,1} \geq n_{k,2} \geq \dots \geq n_{k,q}$ . For  $j < q$ , we have  $g_{k,j} = n_{k,j} - n_{k,j+1}$  units who have exactly  $j$  measurements observed and  $q - j$  measurements missing, and  $g_{k,q} = n_{k,q}$  are the number of subjects with no missing data. At the design stage, the exact rate of dropout will not be known and so  $n_{k,j}$  and thus  $g_{k,j}$  are treated as random variables. In what follows, we denote the probability of having a response observed on a subject in group  $k$  at time point  $j$  by  $p_{k,\text{obs}}(t_{kj}, \delta_k)$ , where  $t_{kj}$  and  $\delta_k$  are the  $j$ th follow-up time point and treatment dose of group  $k$ . The dropout process is noninformative in the sense that it depends on, for example, which treatment the subject receives, and at which time point, but not on any unobserved quantity. Denote  $E[g_{k,j}]$  by  $m_{k,j}$ , then

$$E[g_{k,j}] = m_{k,j} = \begin{cases} Nw_k p_{k,\text{obs}}(t_{kj}, \delta_k), & \text{if } j = q, \\ Nw_k p_{k,\text{obs}}(t_{kj}, \delta_k) - Nw_k p_{k,\text{obs}}(t_{k,j+1}, \delta_k), & \text{if } j < q, \end{cases} \quad (2)$$

where  $Nw_k$  is the number of subjects originally allocated to group  $k$ . The probabilities  $\{p_{k,\text{obs}}(t_{k1}, \delta_k) - p_{k,\text{obs}}(t_{k2}, \delta_k), \dots, p_{k,\text{obs}}(t_{kq-1}, \delta_k) - p_{k,\text{obs}}(t_{kq}, \delta_k), p_{k,\text{obs}}(t_{kq}, \delta_k)\}$  could be considered as the event probabilities of a multinomial distribution.

## 2.3 | Optimal design framework

In the absence of missing responses, a design framework for the linear mixed model constructs an optimal design by finding the setting of design matrix  $\mathbf{X}_k$  such that a function of the Fisher information defined in (1) is optimized over a

design region. Considering the available subject data used in available case analysis, ie, all observed responses of subjects who may or may not dropout at later time points, and the impact of dropouts, the Fisher information can be written as

$$\sum_{k=1}^c \sum_{j=1}^q g_{k,j} \mathbf{X}_{k[j]}^T \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]},$$

where  $\mathbf{X}_{k[j]}$  denotes the subdesign matrix of a subject in group  $k$  measured up to the  $j$ th time point. As the Fisher information is inversely proportional to the variance-covariance matrix of the model parameters, it is common to maximize a function of the Fisher information that corresponds to minimizing the corresponding function of the variance-covariance matrix. A commonly used criterion is to maximize/minimize the determinant of the information/variance-covariance matrix that is typically referred to as  $D$ -optimality. However, at the planning stage of the experiment, the observed values of  $g_{k,j}$ ,  $k = 1, \dots, c, j = 1, \dots, q$ , are not available. So, we aim to maximize a function of the expected information matrix instead, ie,

$$E \left[ \sum_{k=1}^c \sum_{j=1}^q g_{k,j} \mathbf{X}_{k[j]}^T \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]} \right] = \sum_{k=1}^c \sum_{j=1}^q m_{k,j} \mathbf{X}_{k[j]}^T \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]}.$$

We note that taking the inverse of this matrix does not give the expected variance-covariance matrix. However, Galbraith et al<sup>1</sup> found that, in a single cohort study with equally spaced measurements, this simple approximation gives similar results as more complicated approximations to the variance-covariance matrix. For moderate to large sample sizes, this was confirmed by Lee et al<sup>9</sup> for studies without repeated measurements. Also, using this simple approximation in a single cohort longitudinal study, Ortega-Azurduy et al<sup>5</sup> investigated the loss in efficiency of  $D$ -optimal designs that were found assuming the complete data set would be observed, when in fact, dropouts occurred. Hence, we will also use the inverse of the expected Fisher information to approximate the variance-covariance matrix.

Here, we propose a more comprehensive design framework for studies with repeated measurements where more than one group of experimental units are considered in the experiment. This is a common occurrence in most clinical studies where there will be two or more groups followed up, eg, placebo and treatment. Sections 2.3.1 and 2.3.2 derive the relevant framework.

### 2.3.1 | Optimal designs when baseline measurements are comparable

In this section, we assume that the experimental units are recruited from a homogeneous population where different groups of units have comparable baseline measurements at the onset of the study. An example of this type of study is to investigate the efficacy of different treatments over time on subjects who have the same health status. We consider a special case of the linear mixed model where a group indicator matrix,  $\mathbf{K}_i$ , is incorporated in the model formulation, giving

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{K}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i,$$

as the repeated measurements of subject  $i$ . In what follows, we refer to this model as  $\mathbf{M}_g$ . Schmelter<sup>13</sup> uses similar models to find optimal designs for linear mixed models but does not take dropout into account.

To fix ideas, if there are two groups in an experiment, corresponding to, eg, placebo and an active treatment, and the regression function is linear in time,  $\mathbf{M}_g$  has  $\mathbf{K}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$  if experimental unit  $i$  is in group 1; otherwise,  $\mathbf{K}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ , giving for the observation at the  $j$ th time point

$$(\mathbf{X}_i \mathbf{K}_i \boldsymbol{\beta})_j = \begin{cases} \beta_0 + t_{1j} \beta_1, & \text{if experimental unit } i \text{ is in group 1,} \\ \beta_0 + t_{2j} \beta_2, & \text{if experimental unit } i \text{ is in group 2,} \end{cases}$$

with random effects  $(\mathbf{Z}_i \mathbf{b}_i)_j = b_{0i} + t_{kj} b_{1i}$ ,  $k = 1, 2$ . The slope parameters  $\beta_1$  and  $\beta_2$  reflect the effect on responses of group 1 and group 2, respectively, due to a unit change in time;  $(\mathbf{Z}_i \mathbf{b}_i)_j$  reflects the response variability of subjects in different groups at time point  $j$ .

To construct an optimal design for this special case of the linear mixed model, we seek an optimal design

$$\xi^* = \left\{ \begin{matrix} \mathbf{t}'_1 & \mathbf{t}'_2 & \cdots & \mathbf{t}'_c \\ w_1 & w_2 & \cdots & w_c \end{matrix} \right\} \quad (3)$$

that optimizes a function of

$$\sum_{k=1}^c \sum_{j=1}^q m_{k,j} \mathbf{K}_k^T \mathbf{X}_{k[j]}^T \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]} \mathbf{K}_k \quad (4)$$

over the design region of possible time points,  $\mathcal{X}$ , and the weights  $w_k$ , where  $\mathbf{t}'_k = \{t_{k1}, t_{k2}, \dots, t_{kq}\}$  is the optimal allocation of unique time points for measuring an outcome variable on group  $k$ ,  $k = 1, \dots, c$ , and  $m_{k,j}$ , a function of  $w_k$ , is obtained from (2). The optimization is subject to the constraints

$$t_{k1} < t_{k2} < \dots < t_{kq} \quad (5)$$

and

$$0 \leq w_k \leq 1, \quad k = 1, \dots, c, \quad \sum_{k=1}^c w_k = 1. \quad (6)$$

Note that in order to implement such a design, the weights  $w_k$ ,  $k = 1, \dots, c$ , may need to be rounded such that  $Nw_k$  is an integer for all  $k = 1, \dots, c$ . The more relaxed condition (6) is commonly used in the optimal design literature to facilitate numerical design search. A design found under (6) is referred to as an approximate design, whereas a design where all  $Nw_k$ 's are integers is called an exact design.

### 2.3.2 | Extension to incorporate additional design variables

In some scenarios, the experimental conditions of different groups are reflected by a continuous explanatory variable  $\delta$  whose levels can be set by the experimenter. Consider, for example, a treatment that is a dose of a new drug. Although in practice, each participant in the study could have a unique dose level, typically the experimenter would assign individuals to  $c$  distinct values of the dose of the drug including placebo (ie, when the dose equals 0). We assume that the values of the dose can be selected from within some dose range, eg, between placebo and maximum tolerated dose. In this situation, the variable “dose” becomes part of the design, and it may be possible to increase the amount of information to be gleaned from the data by an efficient selection of the doses as well as the time points. Note that we can also treat all individuals receiving the same value of a dose as belonging to a group and allow each group's set of optimal time points to differ, as described in the previous section. In addition, the probability of observing a response at a given time point may also depend on the dose the patient received. In this section, we extend our design framework to incorporate this scenario.

In the simplest example, the  $j$ th repeated measurement of subject  $i$  who is assigned a dose value of  $\delta_k$ , where  $k \in \{1, \dots, c\}$ , is

$$y_{ij} = \beta_0 + t_{kj}\beta_1 + \delta_k\beta_2 + b_{0i} + t_{kj}b_{1i} + \epsilon_{ij},$$

where  $\beta_0$  is the intercept,  $\beta_1$  is the effect on responses of subjects due to a unit change in time,  $\beta_2$  is the effect on responses due to a unit change in dose, and  $\{b_{0i}, b_{1i}\}$  are random effects. We refer to this model as  $\mathbf{M}_d$ . This model can easily be extended to include, for example, a dose-time interaction effect to incorporate the possibility that different doses may affect the responses differently over time, but we will consider  $\mathbf{M}_d$  for illustrative purposes in what follows.

Denote the design matrix of this linear mixed model by  $\mathbf{X}_k(t, \delta)$ . A design problem is then to find

$$\xi^* = \left\{ \begin{pmatrix} \mathbf{t}'_1, \delta_1 \\ w_1 \end{pmatrix}, \begin{pmatrix} \mathbf{t}'_2, \delta_2 \\ w_2 \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{t}'_c, \delta_c \\ w_c \end{pmatrix} \right\}, \quad (7)$$

where the elements,  $\delta_k$ , reflect the experimental conditions of group  $k$ ,  $k = 1, \dots, c$ , such that a function of the matrix

$$\sum_{k=1}^c \sum_{j=1}^q m_{k,j} \mathbf{X}_{k[j]}^T(t, \delta) \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]}(t, \delta) \quad (8)$$

is optimized over the design regions, ie, the design region of time points,  $\mathcal{X}$ , and  $w_k$ , the weight for group  $k$ , as well as the design region of  $\delta$ . Similar to the model in Section 2.3.1,  $m_{k,j}$  is obtained from (2), and this design problem is subject to constraints (5) and (6).

## 2.4 | Locally optimal designs

Having chosen the number of groups,  $c$ , the number of repeated measurements,  $q$ , and the structure of  $\mathbf{V}_i$ , the covariance matrix of  $\mathbf{y}_i$ , for a chosen formulation of the linear mixed model, a design problem is to find the time points of measuring an outcome variable on the groups and the proportion of units to allocate to each group, such that a function of the corresponding information matrix is optimized over the design region  $\mathcal{X}$  (and the design region of  $\delta$  if  $\delta$  is also to be optimized in the design problem of the model introduced in Section 2.3.2). In practice, the structure of  $\mathbf{V}_i$  is not known at the design stage of an experiment. To construct an optimal design for a future experiment, we employ the notion of locally optimal designs. The structure of  $\mathbf{V}_i$  can be estimated using some historical data or the information that is obtained from



some pilot studies. Moreover, the experimenters need to specify some MAR mechanisms for the different groups prior to finding an optimal design for a study with follow-up/repeated measurements.

In addition to the assumption that experimental units are identical and independently distributed, we assume that  $\epsilon_i$  and  $\mathbf{b}_i$  are independent and that a first-order autoregressive process, AR(1), is chosen for  $\epsilon_i$  to capture serial correlation. The AR(1) process has parameter  $0 < \rho < 1$  and covariance structure with elements  $\psi(t_j, t_{j'}) = \rho^{|t_j - t_{j'}|}$ . This process is often used to model time series for experiments where observations measured closer together in time are more correlated than those measured further apart. On the other hand, the random effects  $\mathbf{b}_i$  reflect the between person variation, ie, how individuals behave distinctly in the population. By fixing the covariance matrix of  $\mathbf{b}_i$ , we can find locally optimal group designs for different classes of the above described linear mixed models.

For example, consider a linear regression model with a random intercept and slope denoted by  $\mathbf{b}_i = (b_{0i}, b_{1i})^T$  with covariance matrix

$$\mathbf{D} = \begin{pmatrix} \text{var}(b_{0i}) & \text{cov}(b_{0i}, b_{1i}) \\ \text{cov}(b_{0i}, b_{1i}) & \text{var}(b_{1i}) \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix},$$

a fixed effects model has a zero matrix for  $\mathbf{D}$ ; a random intercept model has  $d_{11} > 0, d_{22} = d_{12} = 0$ ; a random intercept and slope model has  $d_{11} > 0, d_{22} > 0, d_{12} = 0$ ; and a correlated random intercept and slope model has  $d_{11}, d_{22} > 0, d_{12} \neq 0$ .

Together with the fixed effects,  $\mathbf{X}_i \mathbf{K}_i \boldsymbol{\beta}$  for model  $\mathbf{M}_g$  that has the same baseline measurements for different groups, or  $\mathbf{X}_i(t, \delta) \boldsymbol{\beta}$  for model  $\mathbf{M}_d$  where groups additionally depend on a continuous design variable, we can consider the profile of locally optimal time points to measure units' responses for a range of different  $\rho \in (0, 1)$  by optimizing a function of the corresponding information matrix. We note that, often, a numerical optimization procedure is required to find a solution to the design problem.

### 3 | OPTIMAL DESIGNS IN THE PRESENCE OF DROPOUTS

The novelty of this work focuses on two aspects when finding a design for a future experiment/study. Firstly, for a study that employs model  $\mathbf{M}_g$ , we propose to consider two scenarios when finding an optimal design: (1) all experimental units are restricted to have the same set of time points of measuring the outcome variable and (2) different groups are allowed to have different sets of time points of measuring the outcome variable. These can be done by restricting all  $\mathbf{t}'_k$  to be one set of time points in the design problem for scenario (1) and by setting all sets of  $\mathbf{t}'_k$  as free variables to be searched in the optimization problem for scenario (2). In other words, the restricted design condition forces the time points in  $\mathbf{X}_i \mathbf{K}_i \boldsymbol{\beta}$  to be the same for all experimental units whereas the flexible design condition allows the time points in these design matrices to be different for different groups. The former setting is often used in a blinded trial or longitudinal observational study whereas the latter condition could be implemented in an open label/unblinded trial where clinicians and experimental subjects know the administered treatment. Secondly, we propose to optimize the treatment dose levels for a trial that employs model  $\mathbf{M}_d$  when the clinicians have the freedom to do so. In this case, we consider the restricted design condition where all subjects are measured at the same set of time points to conform with common practice. We note that model  $\mathbf{M}_g$  would not have the flexibility for optimizing the dose levels from a design perspective when the dropout mechanism depends on both the time point of follow-up measurements and treatment dose level.

In this section, we illustrate these aspects by considering an experiment that has  $c = 2$  and  $q = 4$  in each of the two groups and that both special cases of the linear mixed model have fixed effects  $\{\beta_0, \beta_1, \beta_2\}$  and random effects  $\{b_{0i}, b_{1i}\}$ . Note that the interpretation of the fixed effect parameters for both types of models is different and that the class of linear mixed model is determined by the values of the covariance matrix  $\mathbf{D}$ . For each class of the linear mixed model, we employed  $d_{11} = 1, d_{22} = 3, d_{12} = 0.8\sqrt{3}$ , respectively, unless the model required some (or all of these) to be set to 0, to study the profile of the optimal time points across a range of  $\rho$  from 0 to 0.9 in step sizes of 0.1. These values are chosen for ease of comparison with the literature.<sup>5</sup> The results presented here are found by using the function *fmincon* in MATLAB, whereby different sets of initial values are employed to verify the locally optimal design.

#### 3.1 | Designs with time dependent dropouts

We now present the application of our design framework for model  $\mathbf{M}_g$ . An example of a design problem is to find  $\mathbf{t}'_1 = (t_{11}, t_{12}, t_{13}, t_{14})$ ,  $\mathbf{t}'_2 = (t_{21}, t_{22}, t_{23}, t_{24})$ , and  $w_1$ , such that the determinant of (4) is maximized over the design region of time points. The design problem is subject to constraints (2), (5), (6) and the flexible/restricted design condition on the time

points. Note that maximizing the determinant of (4) is equivalent to minimizing the determinant of the covariance matrix for the estimator of the population parameter vector  $\hat{\beta}$  and is defined as  $D$ -optimality.<sup>21</sup>

For better comparability of our results with the literature that did not account for the presence of dropouts,<sup>5,11</sup> and the work that investigated the efficiency loss of  $D$ -optimal designs due to the presence of dropouts in a longitudinal study with one cohort,<sup>5</sup> we consider the standardized time design region  $[-1, 1]$  that reflects the interval between follow-up visits in this illustration. We consider  $\sigma^2 = 1$ , a linear response probability function

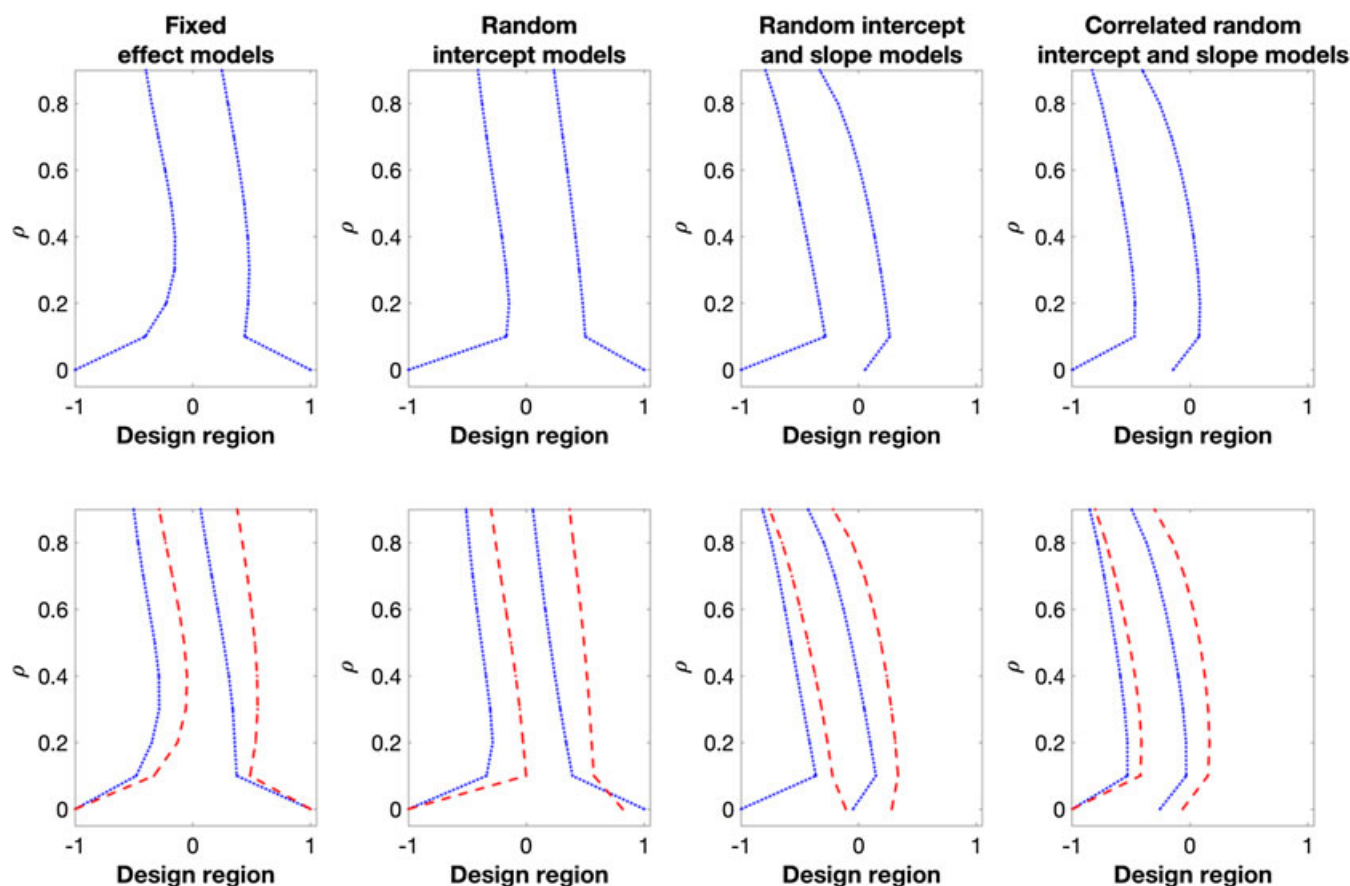
$$p_{1,\text{obs}}(t, \delta_1) = 0.65 - 0.35t, \quad (9)$$

and a quadratic response probability function

$$p_{2,\text{obs}}(t, \delta_2) = 0.5 - 0.35t + 0.15t^2, \quad (10)$$

to capture the presence of missing responses in a group. With this missing mechanism, subjects in group 1 are more likely to be observed for longer than subjects in group 2. We note that our design framework is compatible with a wide class of noninformative dropout mechanisms that have monotone response probability functions. We conjecture that Ortega-Azurduy et al<sup>5</sup> considered these less commonly used response models for ease of satisfying constraint (2) in the design framework, whereby the sum of event probabilities must equal to one. When a response model that does not satisfy this constraint over the time design region is used in our design framework, we resolve this conflict by not optimizing the first two time points whereby they must be chosen by the experimenter, and truncate the time design region to find  $t_{k3}, \dots, t_{kq-1}$ . See Section 3.2 for illustration.

To find the optimal time points, we choose the lower bound of the time design region as the first time point for both groups, ie,  $t_{11} = t_{21} = -1$  in this example, because it has the highest response probability rate over the region; and choose the upper bound as the last time point, ie,  $t_{14} = t_{24} = 1$ , for practical reasons (eg, pre-selected end of study time). The



**FIGURE 1** The middle two  $D$ -optimal time points for model  $M_g$  with  $c = 2$ ,  $q = 4$ , restricted design condition (top row) and flexible design condition (bottom row), respectively. In the bottom plots, Group 1 (blue dotted lines) has quadratic response probability function (10); Group 2 (red dashed lines) has linear response probability function (9) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**TABLE 1** Maximum/minimum optimal weight,  $w_1$ , for the group with quadratic response, found under the restricted and the flexible design condition, respectively, for model  $M_g$  (corresponds to the optimal designs in Figure 1)

	Flexible		Restricted	
	Max $w_1$	Min $w_1$	Max $w_1$	Min $w_1$
FE	0.5000 <sup>a</sup>	0.4821	0.5000 <sup>a</sup>	0.4828
RI	0.4981	0.4901	0.5000 <sup>a</sup>	0.4878
RIRS	0.4921	0.4624	0.4921	0.4781
RIRSc	0.4907	0.4761	0.4907	0.4773

Abbreviations: FE, fixed effects model; RI, random intercept model; RIRS, random intercept and slope model; RIRSc, correlated random intercept and slope model.

<sup>a</sup>These maximum weights are obtained for  $\rho = 0$ .

middle time points of measuring the outcome variable on the groups and the corresponding weights can be found by using optimization algorithms.

The  $x$ -axis on the plots in Figure 1 shows the second and the third time points of locally  $D$ -optimal designs for various versions of model  $M_g$ , with quadratic response probability function (10) in one group, and linear response probability function (9) in the other group. Each plot in Figure 1 corresponds to the middle time points of the optimal designs for each class of the models; the pair of dotted lines in the first row of plots correspond to the second and third time points assuming both groups are measured at the same set of time points (scenario 1); the two pairs of lines in the second row of plots correspond to the sets of time points of measuring the outcome variable in each group, which are found assuming these can differ between groups (scenario 2). The  $y$ -axis in each plot shows the considered value of  $\rho$  (with 0.1 between each case) in the design problems.

Looking at the trend of the time points, we find that the optimal time points for the experiments with the two response probability functions do not converge to the equidistant design as  $\rho$  approaches to one, in particular for the model with independent random intercept and slope parameters, and the model with correlated random intercept and slope parameters. To be more specific, consider the second row of plots in Figure 1, we find that the optimal time points of measuring the outcome variable on the experimental units who remain longer in the study, ie, the group with the linear response probability function, are larger than those of the subjects who are expected to be dropping out earlier from the study, ie, the group with the quadratic response probability function. Intuitively, this is reasonable as we would like to observe units' follow-up measurements before they drop out. Comparing the first and second rows of plots, we learn that there may be considerable differences between the optimal time points in different groups that are found under the respective design conditions. Hence, when planning a study in practice, it should be investigated if using different time points for different groups is feasible (eg, would not violate double-blindness) since it would increase the amount of information that can be gathered from the data.

We now consider the weights of the locally  $D$ -optimal designs for the above described experiment, ie, the proportions of experimental units allocated to each group. Table 1 shows the maximum and minimum weights across the 10 considered cases of  $\rho$ ,  $0 \leq \rho \leq 0.9$  (and difference of 0.1 between each case). For the experiments with  $\rho > 0$ , the design framework maximizes the expected total information by having more experimental units in the group that has a higher response rate within the time region, ie, the group that has the linear response probability function (red dashed line) in the plots in Figure 1, respectively. On the other hand, for the experiment with  $\rho = 0$  in scenario 1, the locally  $D$ -optimal designs for the corresponding fixed effect models and the random intercept models have  $w_1 = 0.5 = w_2$ . This is because all the experimental units have the same response rate at the third/fourth optimal time points, ie, at the end points of  $\mathcal{X}$  (see the first two plots from the left in the first row of plots in Figure 1). The same reason also applies to the locally  $D$ -optimal design for model  $M_g$  with fixed effect parameters and  $\rho = 0$  in scenario 2 (see the first plot from the left in the second row of plots in Figure 1).

### 3.2 | Designs with time and dose dependent dropouts

We now present the application of our design framework for model  $M_d$  where we have the flexibility to additionally choose the dose levels, which we treat as comparative groups in the trial, while also simultaneously assessing the effect of dose

on the outcome. We restrict  $t'_2 = t'_1$  in this illustration for ease of presentation and consider a response model

$$p_{\text{obs}}(t, \delta) = \frac{1}{1 + \exp(\gamma_0 + \gamma_1 \delta + \gamma_2 t)},$$

with  $\gamma_1 < 0$  and  $\gamma_2 > 0$ , so that the response rate of subjects increases with dose level and decreases with time. In the situation where the response rate decreases with dose level, we could choose  $\gamma_1 > 0$  and  $\gamma_2 > 0$  to reflect the elicitation of the response rate. Furthermore, we could also include a dose-time interaction in the linear predictor above. The logistic function, while commonly used in the literature to model missingness of responses, would never lead to the event probabilities of the multinomial distribution in constraint (2) summing to one or, equivalently, would not result in observing a response at time zero (baseline) with probability one. We resolve this issue by fixing the second time point of measurements (in addition to the first measurement at time zero), and truncate the design region to optimize the remaining points  $t_{k3}, \dots, t_{kq}$ , without normalizing them after an optimal design is found.

As an illustration, consider a one-year trial that examines the effect of a drug to treat Alzheimer's disease.<sup>14</sup> Patients are measured five times during the year, with the first measurement taken at baseline, ie, the start of the trial. We assume without loss of generality that the dose levels could take values within a range of 0 (placebo) up to 100. We would choose  $t_{11} = 0$  (corresponds to baseline measurement at day 0) and  $t_{15} = 364$  (corresponds to the end of the study), and then fix, for example,  $t_{12} = 42$  for the first follow-up visit after 1/4 year. The design problem is then to find  $t'_1 = (t_{13}, t_{14})$  in design region  $(42, 364)$ , and treatment dose levels  $\delta_1$  and  $\delta_2$  in  $[0, 100]$ , such that the determinant of (8) is maximized over this design region, subject to constraints (2) and (5). If the weight,  $w_1$ , is not fixed to 0.5 in advance, the design problem would also involve finding  $w_1$  and  $w_2$  subject to constraint (6). Consider  $\sigma^2 = 10$ ,  $\{\gamma_0, \gamma_1, \gamma_2\} = \{0, -3/100, 3/364\}$  and the corresponding values for matrix  $D$  that determine the respective classes of linear mixed models. We find optimal designs by choosing  $t_{15} = 364$  to reflect the pre-selected end of trial time, and  $\delta_2 = 100$  as it has the highest response probability rate over the dose region. We find the middle time points of follow-up measurements,  $\delta_1$  and  $w_1$  (if it was not pre-chosen) using *fmincon* in MATLAB.

Figure 2 shows the  $D$ -optimal settings for the above scenario on the  $x$ -axis of the plot. The first row of plots correspond to the middle time points of the  $D$ -optimal designs for a range of  $\rho$ ; the second row of plots corresponds to the optimal  $\delta_1$ ; the third row of plots corresponds to  $w_1$  for the case in which  $w_1$  is not pre-selected in the design problem. In all plots, blue solid lines correspond to the optimal design setting that has optimized  $w_1$ ; red dashed lines correspond to the design problem where  $w_1 = 0.5$ . The  $y$ -axis in each plot shows the value of  $\rho$  (in steps of 0.1) in the design problems.

Looking at the first row of plots, we find that the optimal time points are the same for both designs assuming equal or unequal weights. Moreover, the optimal time points are far from an equal time interval design that would measure subjects at approximately  $\{42, 149, 256, 364\}$  days after baseline measurements. If  $w_1$  is not pre-chosen in the design problem, we find that a  $D$ -optimal design has  $w_1 < 0.5$  and smaller optimal  $\delta_1$  when compared to  $\delta_1$  of the optimal design that has  $w_1 = 0.5$ . The former finding agrees with the finding for model  $M_g$  that, in the presence of dropouts, having more subjects from the group that has a higher response rate would increase the precision of the fixed effect parameters estimation. If we insist in having equal sample size for both groups, having a larger dose than 0 would hopefully decrease the overall dropout rate of that group.

## 4 | APPLICATION: REDESIGNING A STUDY ON ALZHEIMER'S DISEASE

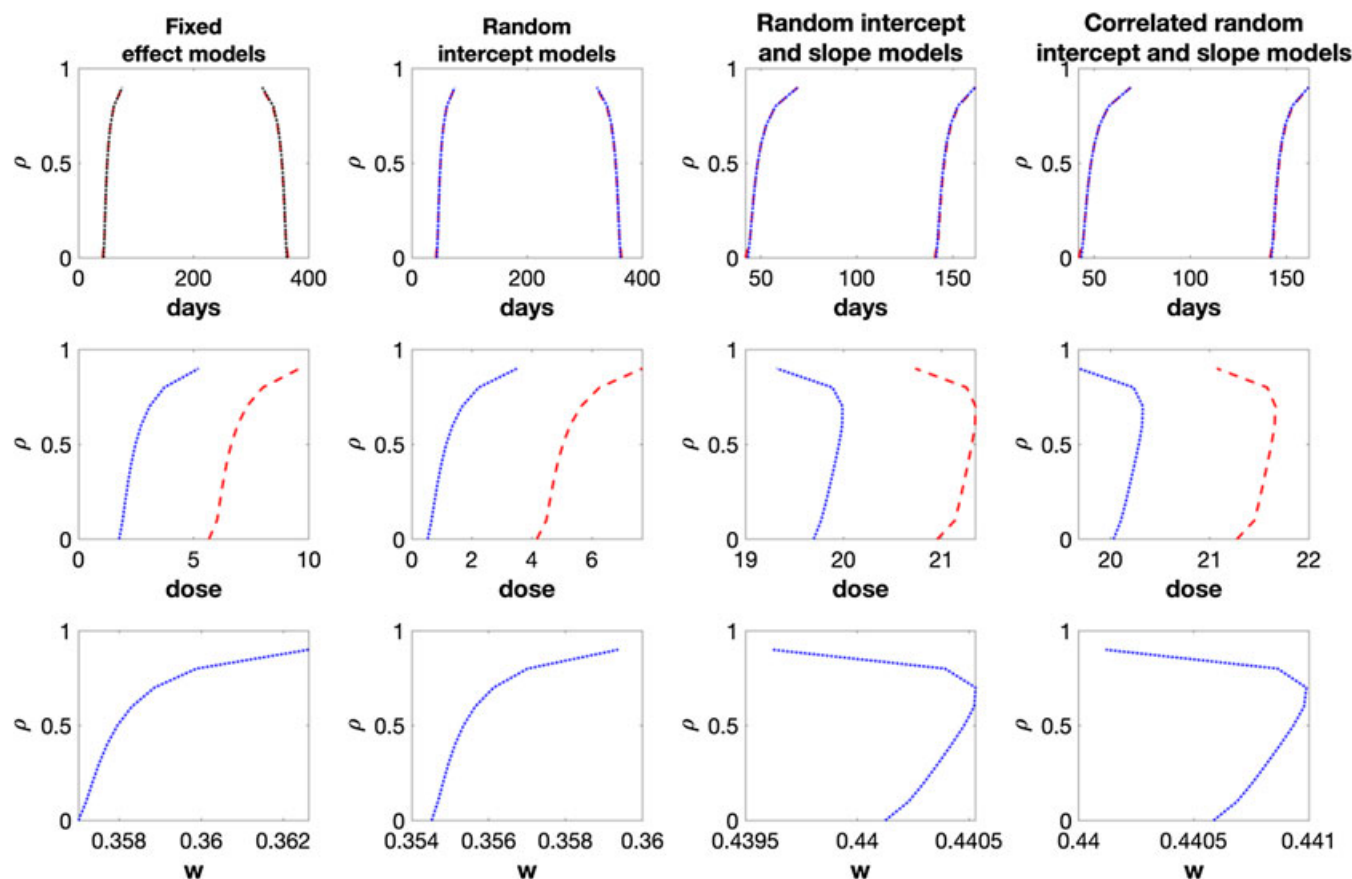
### 4.1 | Background and exploratory analysis

We now illustrate our framework through an application using the data from the Alzheimer's disease study.<sup>14</sup> The study<sup>14</sup> considered the effectiveness of two drugs to treat Alzheimer's disease, donepezil and memantine, as well as receiving both drugs. The study randomized patients into four groups given by a factorial design, with each group of patients followed up at four subsequent times after baseline. Each patient had five measurements in each group at week 0, 6, 18, 30, and 52, respectively. The aim of the study is to explore the changes from baseline measurement in each group over a period of 52 weeks. For illustration purposes, we only consider the experimental units in the placebo group and the donepezil-memantine (treatment) group, who were included in the primary intention-to-treat sample. Here, we treat the primary outcome measure, SMMSE score (higher score indicates better cognitive function), as the response variable of our model. The total sample size of the data used in this illustration is  $N = 144$  (72 in each group). By the end of the study period, there were only 29 patients in the placebo group and 51 patients in the treatment group included in the

per-protocol analysis, with some of them having been lost to follow-up during the course of the study. We consider time in terms of days for both the response probability and the design. The original design,  $\xi_{\text{ori}}$ , measures subjects at day 42, 126, 210, and 364 after the baseline measurement.

To illustrate our proposed framework, we assume an analyst is also interested in investigating the effect of the specific dose level prescribed to patients on their cognitive function, as well as studying how their cognitive function changes over time in the different dose groups. This results in adopting model  $M_d$  as our analysis model and finding an optimal design accordingly. We note that the simpler model  $M_g$  could be considered here as well, but this would not allow any optimal dose allocation to be considered, and could potentially lead to a loss of information in learning about the dose-response relationship. To find an optimal design, treating dose as a design variable to be optimized, we assume the design will have two dose levels/groups. In the original design, this corresponds to the placebo and the treatment group. Without loss of generality, we assume that the dose can take values on the interval  $[0, 100]$ . We see from the data that the treatment group has less missing data than the placebo group. So, we set the upper optimal dose level to be  $\delta_2 = 100$ , as was also the case in the previous illustration. The optimal design then proceeds to find the optimal value of the lower dose level,  $\delta_1$ , as well as the weight to assign to each dose group, and the optimal time points at which to take measurements over the one-year period. We assume initially that measurements will be taken at five time points over the one-year period (baseline plus four further follow-up measurements) and assume that both dose groups will be observed at the same set of time points. Furthermore, we assume the last measurement will be taken at the end of the trial period (one year) and that the first follow-up measurement should not be taken too close to the baseline measurement to give the treatment some time to (potentially) start showing an effect. In particular, we set the time range (in days), in which the three middle time points must lie, to  $[42, 364]$ .

As designs are locally optimal, we first use the *lme* function in the *R* software package to fit the SMMSE score to the two groups' available data ( $n_1 = 29$  in placebo group and  $n_2 = 51$  in treatment group), to obtain a realistic scenario for which



**FIGURE 2** First row of plots: middle two  $D$ -optimal time points for model  $M_d$ ; second row: optimal dose  $\delta_1$ ; third row: optimal weight  $w_1$ . The blue solid lines correspond to the optimal designs that involve choosing  $w_1$ ; the red dashed lines correspond to the optimal designs that have pre-chosen  $w_1 = 0.5$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

to design. We consider all four possible classes of model  $M_d$  and three different temporal correlation structures within each subject: no correlation, compound symmetry where  $\psi(t_j, t_{j'}) = \rho$  for  $t_j \neq t_{j'}$ , and AR(1) as defined in Section 2.4. Table 2 presents Akaike information criterion (AIC) and Bayesian information criterion (BIC) values from the different models fitted to the data. We find that the random intercept model with AR(1) correlation and parameter estimates  $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\} = \{8.939, -0.0866, 0.0146\}$ ,  $\{\hat{\rho}, \hat{d}_{11}, \hat{d}_{22}, \hat{d}_{12}\} = \{0.3326, 2.661^2, 0, 0\}$ , and  $\hat{\sigma}^2 = 2.613^2$  has the smallest AIC and BIC values among the possible classes of models. To obtain realistic response probability functions for the two groups, we use the numbers of subjects who remain in the study over the period of 364 days<sup>14</sup> to fit logistic regression models for the two groups where the placebo group has  $\delta = 0$  whereas the treatment group has  $\delta = 100$ . We obtain the following response probability:

$$\hat{p}_{\text{obs}}(t_{ij}, \delta) = \frac{1}{1 + \exp(\hat{\gamma}_0 + \hat{\gamma}_1 \delta + \hat{\gamma}_2 t_{ij})},$$

with  $\{\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2\} = \{-2.2332, -0.0131, 0.0100\}$ , reflecting that the treatment group has a larger response rate than the placebo group. These values can then be used to construct optimal designs. The type of fixed and random effects models we have considered here are commonly used to analyze repeated measurement clinical trial data where there are likely to be dependencies within an individual's measurements collected over time.<sup>5</sup> In addition, the logistic model is a commonly used model to characterize missing data mechanisms<sup>18,22,23</sup> and is a convenient model to use when seeking to fit a regression model for binary response data.

## 4.2 | Optimal designs, simulation study, and sensitivity analysis

We now determine an optimal design using the  $D$ -optimality criterion for model  $M_d$  assuming a random intercept with parameter values specified above. We denote this design by  $\xi_D^*$  and it is presented in Table 3 together with the expected number of subjects who have  $j$  observations/measurements,  $m_{\delta,j}$ , where  $j = 1, \dots, 5$ . The design would measure all subjects at 42, 285, 356, 364 days, substantially different to time points used in the original design, denoted by  $\xi_{\text{ori}}$  in Table 3. The design would also allocate  $n_1 = 144 * 0.42 \approx 60$  subjects in the placebo group and  $n_2 = 144 - 60 = 84$  subjects in the treatment group, again substantially different to the allocation used in  $\xi_{\text{ori}}$ . The design  $\xi_D^*$  determines  $\delta_1 = 0$ , ie,  $\xi_D^*$  takes the bounds of the dose design region as the optimal values. This is because a value of  $\gamma_1$  close to zero was taken when constructing the design and so the optimal design chooses the bounds of the dose design region in order to maximize the determinant of (8). Nevertheless, in other situations, other dose levels may be determined to be optimal, or it may be the case that the practitioner would like to specify more than two dose levels for the design, eg, if they think there is a more complex dose response relationship, which should then also be reflected in the choice of model.

In practice, the correct model for the data and true parameter values will not be known in advance of constructing optimal designs. To address this, we perform a sensitivity analysis by considering alternative locally optimal designs based on different model formulations and parameter values used to construct  $\xi_D^*$ . Table 3 presents designs  $\xi_{D,RI}^*$ ,  $\xi_{D,RIS}^*$ , and  $\xi_{D,RISC}^*$ , which are  $D$ -optimal for the values  $\{\rho, d_{11}, d_{22}, d_{12}\} = \{0.3326 \times 2, (2.661 \times 2)^2, 0, 0\}$ ,  $\{0.3326, 2.661^2, 2, 0\}$ , and  $\{0.3326, 2.661^2, 2, -1\}$ , respectively. In other words, these optimal designs correspond to a random intercept model, a linear mixed effect model with uncorrelated random intercept and slope parameters, and a linear mixed effect model where the random parameters are correlated, but with parameter estimates different to those obtained from the data. In particular, note that design  $\xi_{D,RI}^*$  has the same model formulation as  $\xi_D^*$  but is constructed using a different set of parameters. A similar sensitivity analysis could be done when the dropout mechanism is uncertain, although we do not present the investigation here.

We see that  $\xi_{D,RI}^*$  would measure subjects at 42, 292, 349, and 364 days after baseline measurement that is very similar to the optimal follow-up times resulting from  $\xi_D^*$ . The optimal allocation to the two dose groups is also the same under  $\xi_{D,RI}^*$  as  $\xi_D^*$ . Both these optimal designs have  $t_{13}$  and  $t_{14}$  which are very close to the bounds of the time point design region [42, 364], which might suggest that having one fewer time point might be more efficient. This will be explored more in the

**TABLE 2** Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for the different models fitted to the data in Section 4.1. Compound symmetry (CS) and AR(1) refer to the serial correlation structure within subjects over time

	FE	FE CS	FE AR(1)	RI	RI CS	RI AR(1)	RIRS	RIRS CS	RIRS AR(1)
AIC	2171.89	1972.30	1969.32	1972.30	1974.30	1961.31	1967.65	1969.65	1962.80
BIC	2187.73	1992.10	1989.12	1992.10	1998.06	1985.07	1991.41	1997.37	1990.53

**TABLE 3** Middle time points of  $D$ -optimal designs for several classes of model  $M_d$ ,  $t_{11} = t_{21} = 0$ ,  $t_{12} = t_{22} = 42$ ,  $t_{15} = t_{25} = 364$ ,  $w_2 = 1 - w_1$ . The optimal values of  $\delta$  are the bounds of the design region  $[0, 100]$ ,  $w_1$  is the optimal weight of the placebo group ( $\delta_1 = 0$ ). Both groups are measured at the same time points,  $m_{\delta=k,j}$ ,  $k = \{0, 100\}$ ,  $j = 1, \dots, 5$  is the expected number of subjects in group  $k$  who have exactly  $j$  observed responses

	$t_{11}$	$t_{12}$	$t_{13}$	$t_{14}$	$t_{15}$	$w_1$
	$m_{\delta=0,1}$	$m_{\delta=0,2}$	$m_{\delta=0,3}$	$m_{\delta=0,4}$	$m_{\delta=0,5}$	$n_1$
	$m_{\delta=100,1}$	$m_{\delta=100,2}$	$m_{\delta=100,3}$	$m_{\delta=100,4}$	$m_{\delta=100,5}$	$n_2$
<b>Five Time Point Design, <math>N = 144</math></b>						
$\xi_{ori}$	0	42	126	210	364	0.5
Placebo group	10	10	14	24	14	72
Treatment group	3	4	7	24	34	72
$\xi_D^*$	0	42	285.2340	355.6943	364	0.4221
Placebo group	8	31	8	1	12	60
Treatment group	4	24	14	2	40	84
$\xi_{D,RI}^*$	0	42	292.2367	349.1291	364	0.4189
Placebo group	8	32	7	1	12	60
Treatment group	3	26	12	3	40	84
$\xi_{D,RIS}^*$	0	42	46.3915	153.7180	364	0.4865
Placebo group	10	0	13	33	14	70
Treatment group	3	0	6	30	35	74
$\xi_{D,RISc}^*$	0	42	46.3841	153.8501	364	0.4865
Placebo group	10	0	13	33	14	70
Treatment group	3	0	6	30	35	74
<b>Four Time Point Design, <math>N = 172</math></b>						
$\xi_{D,4}^*$	0	42	318.5670		364	0.4183
Placebo group	10	42	6		14	72
Treatment group	4	37	11		48	100
$\xi_{D,RI}^*$	0	42	322.3673		364	0.4154
Placebo group	10	42	5		14	71
Treatment group	4	38	11		48	101
$\xi_{D,RIS}^*$	0	42	137.3887		364	0.4865
Placebo group	12	13	43		16	84
Treatment group	4	5	37		42	88
$\xi_{D,RISc}^*$	0	42	136.9573		364	0.4865
Placebo group	12	13	43		16	84
Treatment group	4	5	37		42	88

next section. When a model with random intercept and slope is considered at the design stage of a study, the  $D$ -optimal designs become qualitatively more different from the former designs. From Table 3, we see that  $\xi_{D,RIS}^*$  and  $\xi_{D,RISc}^*$  would measure subjects at 42, 46, 154, 364 days after baseline measurements, with both allocating  $n_1 = 144 * 0.4865 \approx 70$  subjects in the placebo group and  $n_2 = 144 - 70 = 74$  subjects in the treatment group. In this example,  $\xi_{D,RIS}^*$  and  $\xi_{D,RISc}^*$  are similar as the value of  $d_{12} = -1$  is relatively small compared with the variances of the random effects. As with  $\xi_D^*$ , all designs choose the bounds of the dose design region, ie,  $\delta_1$  equals 0 in all cases.

We evaluate the performance of different designs through a simulation study. For a given design, we simulate

$$y_{ij} = 8.939 - 0.0866t_{ij} + 0.01458\delta_i + b_{0i} + \epsilon_{ij},$$

where  $t_{ij}$  corresponds to the  $j$ th time point of measuring the outcome variable on subject  $i$ . Here,  $\delta_i = 0$  corresponds to the placebo group, with size  $n_1$ ;  $\delta_i = 100$  corresponds to the treatment group, with size  $n_2$ ;  $\epsilon_{ij} \sim N(0, \hat{\sigma}^2 \hat{\Psi})$ ,  $\hat{\Psi}(t_j, t_{j'}) = \hat{\rho}^{|t_j - t_{j'}|}$ , and  $b_{0i} \sim N(0, \hat{d}_{11})$ . We have thus used model parameters estimated from the data above and thus design  $\xi_D^*$  should perform best. Missing values in the vector of observed responses of subject  $i$ , ie,  $\mathbf{y}_i = \{y_{i1}, \dots, y_{i5}\}$ , are introduced using a multinomial indicator with the corresponding event probabilities computed by  $\hat{p}_{obs}(t_{ij}, \delta)$  for each group.



**TABLE 4** Simulation output of design comparison

	$\text{var}(\hat{\beta}_0)$ $\times 10^{-1}$	$\text{var}(\hat{\beta}_1)$ $\times 10^{-7}$	$\text{var}(\hat{\beta}_2)$ $\times 10^{-5}$	$ \text{cov}(\hat{\beta}) $ $\times 10^{-12}$	$RE_D(\xi, \xi_D^*)$	$RE_D(\xi, \xi_{D,4}^*)$
$\xi_{\text{ori}}, N = 144$	1.524	18.06	2.104	2.828	0.8140	0.7448
$\xi_D^*, N = 144$	1.736	8.146	2.815	1.526	1	0.9150
$\xi_{D,RI}^*, N = 144$	1.745	8.175	2.829	1.544	0.9959	0.9113
$\xi_{D,RIS}^*, N = 144$	1.385	13.17	2.540	2.099	0.8990	0.8226
$\xi_{D,4}^*, N = 172$	1.477	8.440	2.433	1.169	1.093	1

$\xi_{D,RIS}^*$  and  $\xi_{D,RISC}^*$  are identical in this illustration.

For each of the different designs, we repeatedly simulate the incomplete data 100 000 times as described above. For each incomplete data set, we compute the sample estimates for the fixed effect coefficients,  $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\}$  from the available cases. The elements of  $\text{cov}(\hat{\beta})$  are then estimated empirically using the sample estimates that are obtained from each simulated data set.

Table 4 presents the empirical values of  $\text{var}(\hat{\beta}_0)$ ,  $\text{var}(\hat{\beta}_1)$ ,  $\text{var}(\hat{\beta}_2)$ , and  $|\text{cov}(\hat{\beta})|$  from 100 000 simulated sets, as well as the  $D$ -efficiency relative to  $\xi_D^*$  (five time points), the final column will be explained in the next section. The relative  $D$ -efficiency of a design  $\xi$ , with respect to a design  $\xi^*$  is  $RE_D(\xi, \xi^*) = \left( \frac{|I_{\xi^*}|}{|I_{\xi}|} \right)^{1/p}$ , where  $I_{\xi}$  denotes the covariance matrix provided by using the design  $\xi$ , and  $p$  is the number of fixed effect coefficients in the model.

As designs are obtained using  $D$ -optimality, we focus on  $|\text{cov}(\hat{\beta})|$ . From Table 4, we see that  $\xi_{\text{ori}}$  performs poorly with respect to this criterion. We find that the  $D$ -efficiency of the original design,  $\xi_{\text{ori}}$ , relative to  $\xi_D^*$  is 0.814, implying that approximately five replicates of  $\xi_{\text{ori}}$  would be as efficient as four replicates of the  $D$ -optimal designs,  $\xi_D^*$ . When an optimal design assumes the true structure of  $\mathbf{V}_i$  but wrong values for  $\rho$  and  $d_{11}$ , we find that the efficiency loss is not significant; an example of such a design is  $\xi_{D,RI}^*$  that loses only about 0.4%  $D$ -efficiency relative to  $\xi_D^*$  in the simulation. However, if an optimal design assumed the wrong structure of  $\mathbf{V}_i$ , the  $D$ -efficiency loss could be significant. In our illustration, this loss is about 10% when  $\xi_{D,RIS}^*$  or  $\xi_{D,RISC}^*$  is used in the simulation instead of  $\xi_D^*$ . Note, however, that all the derived designs perform better than the original design, even the ones which have been obtained using a misspecified model structure. Similar analysis could be conducted to study efficiency loss when using an optimal design that assumes different noninformative dropout mechanisms.

### 4.3 | Cost saving design

From Table 3, we can see that the derived designs (ie, designs other than  $\xi_{\text{ori}}$ ) have two follow-up time points quite close together. For example,  $\xi_D^*$  and  $\xi_{D,RI}^*$  have the fourth follow-up time point close to the fifth (end of study follow-up). There is thus evidence to suggest that a design could have fewer follow-up time points without greatly suffering from a loss of information. From a practical point of view, having fewer follow-up time points would reduce the overall cost of the study and thus potentially allow more patients to be recruited. This may also reduce the chances that subjects may be lost to follow-up.

We illustrate the benefit of employing an optimal design with a fixed budget where the trade-off is between the number of time points of measurement and total sample size. Specifically, we assume that it costs approximately twice as much to recruit a patient to the study (and take measurements at baseline) as it does to take a follow-up measurement. The cost function ratio used here is a representative example and was obtained through email communication with one of the authors of the original DOMINO study who compared testing and cost templates for a number of studies they had previously done and also looked at the National Institute for Health Research costing templates. They concluded that despite the variability, the relative cost is approximately 2 to 2.5, ie, to recruit a new patient into a study costs about 2 to 2.5 times the cost of an average follow-up visit (Prof C. Holmes, personal communication, February 20, 2017). Under these conditions, the cost saving afforded from moving from a five-time-point to a four-time-point design (including baseline) allows the possible number of patients to be recruited to the study to increase from 144 to approximately 172. As the costs of recruiting patients and taking follow-up measurements should be known in advance of the trial, we do not consider robustness to misspecification of the cost function here. However, we note that, if cost functions are liable to change during the course of a very long longitudinal study that involves recruiting patients over a substantial period of time, then a sensitivity analysis against different cost functions could be performed and a design chosen accordingly.



The lower panel of Table 3 shows the corresponding  $D$ -optimal designs with four time points, but now based on an overall sample size of 172 patients. In this case,  $\xi_{D,4}^*$  would have  $n_1 = 72$  subjects in the placebo group and  $n_2 = 100$  subjects in the treatment group, with follow-up measurements at 42, 319, and 364 days after baseline measurement. The corresponding designs for  $\xi_{D,RI}^*$ ,  $\xi_{D,RIS}^*$ , and  $\xi_{D,RISC}^*$ , now assuming four time points and a sample size of 172, are also presented here for completeness. As before, all designs select the bounds of the dose region, so  $\delta_1 = 0$ .

We also evaluate the performance of  $\xi_{D,4}^*$  in the simulation study. The final column in Table 4 presents the  $D$ -efficiency of all the designs considered relative to  $\xi_{D,4}^*$ , the optimal design based on four time points and a sample size 172. We see that  $\xi_{D,4}^*$  outperforms all the competing designs. The original design,  $\xi_{ori}$ , would lose about 26% efficiency, the five-time-point optimal designs for a random intercept model,  $\xi_D^*$  and  $\xi_{D,RI}^*$ , would lose about 9% efficiency and  $\xi_{D,RIS}^*$  and  $\xi_{D,RISC}^*$  would lose about 18% efficiency. The loss of information by having one fewer follow-up measurement is more than compensated for by the ability to increase the sample size, which leads to an overall increase in information. The cost ratios used in this analysis are consistent with typical values associated with multiarm trials on Alzheimer's disease. We note that practitioners could implement such analysis for a range of different cost scenarios to find the most suitable optimal design for implementation. This section highlights the important role that optimal design techniques can play in obtaining the most efficient design and hence the most information from the trial under fixed budget constraints.

## 5 | CONCLUSION AND DISCUSSION

We have developed an optimal design framework in the presence of dropouts for a large class of linear mixed models and have illustrated our methodology through assessing the optimal designs for two special formulations of the linear mixed model, model  $M_g$  and model  $M_d$ . In particular, our framework for model  $M_d$  allows a design to be constructed, which maximizes information about both the dose-response relationship as well as the relationship between different dose groups over time, taking into account dropout, a hitherto uninvestigated problem. Our framework provides optimal designs for multiarm studies with follow-ups and allows differential noninformative dropout processes for different groups. Our framework accounts for dropout processes that are dependent on the design variables. This assumption is more realistic since not only time but also the treatment may have an effect on dropout. Current literature<sup>5,11,13</sup> could be viewed as special cases of our framework. Moreover, for model  $M_g$ , we have studied two different experimental conditions, ie, a restricted condition where all experimental subjects must have the same set of optimal time points of measuring the outcome variable, and a flexible condition that allows for having different sets of optimal time points of measuring the outcome variable on different groups. We found that, in some scenarios where the missing mechanisms of different groups vary significantly, the flexible designs could be more efficient. So, we recommend their use in practice unless this is prevented by conflicts with secondary objectives of the study or implementation issues such as double blinding in a clinical trial. We note that this experimental condition could also be considered when finding an optimal design for model  $M_d$ .

We have applied our methodology to a real-world example by redesigning a clinical trial for Alzheimer's disease.<sup>14</sup> To this end, we investigated the design problem from two different angles. First, we generated optimal designs for the exact scenario (five time points) of the trial under consideration and found that, if an optimal design had been employed in the original study, almost 19% of the experimental effort (and thus experimental costs) could have been saved while obtaining the same amount of information from the resulting data. The optimal designs for this scenario had several time points clustered relatively close together, which may be impractical (repeated visits to the clinic at short time intervals). This suggested that we may not need five time points for the trial. Hence, we investigated this problem within a cost-efficiency framework and found that optimal designs with four time points and increased sample size such that the overall cost of the trial is kept fixed can lead to more efficient designs, and thus more information to be gleaned from the data.

We have illustrated our methodology using  $D$ -optimality, as one of the most commonly used optimality criteria, which reflects a situation where there is equal interest in estimating all mean model parameters accurately while treating the variance components as nuisance parameters. In practice, the criterion that best reflects the purpose of the experiment should be chosen. If there is more than one such objective, a compound criterion can be selected. In all these cases, our methodology can be followed analogously, with the obvious adjustments to the numerical design search routine.

We have shown that the presence of dropouts has considerable impact on the locally optimal designs for studies involving multiple comparison groups and several follow-up sessions. Only locally optimal designs are available for the linear mixed models as the dropout processes and the covariance structure of the repeated measurements are unknown at the design stage of an experiment. Nevertheless, locally optimal designs are important as benchmarks, against which all other

candidate designs can be assessed. Using data from historical/pilot studies, we can estimate the covariance structure as well as the dropout processes for constructing an optimal design for a future study. In our investigations, we assume an AR(1) process for the observational errors of the experimental subjects to capture the within-subject correlation. By trying different sets of values for the variance of random coefficients, we find that the structure of  $\mathbf{D}$ , ie, the covariance matrix of random coefficients, rather than the values of its elements has more impact on the trends of the optimal time points across the range of realistic values for the serial correlation parameter  $\rho$ . Section 4 illustrated the sensitivity analysis of different locally optimal designs using simulation. We note that further robustness checks could be performed depending on the context. These include robustness to misspecification of the missing data mechanism, misspecification of the serial correlation structure of the errors in the repeated measurements, and misspecification of the covariance structure of random coefficients. If sensitivity analyses flag up a lack of robustness to different parameter values or models, a way to address this could be through an extension of our framework to Bayesian<sup>24</sup> designs or maximin efficient<sup>12</sup> designs.

For finding optimal designs, we have combined the concepts of cost functions based on financial considerations and patient drop-out based on time and treatment. It might be interesting to also factor in a willingness to come function, potentially based on the total number and on the frequency of visits. This could be estimated using historical trial data or a questionnaire administered prior to the start of the trial. For example, some patients may find a large number of visits or visits within a few weeks of each other excessive and may be less likely to attend all of them.

It has been our primary intention to propose a flexible framework for trial designs that allows the possibility for different arms to have different designs. From our point of view, blinded trials could be less efficient than the flexible design that has different schedules for different groups, in terms of collecting sufficient observations given the same resources and potentially different dropout mechanisms. In this context, the flexible design can be viewed as a benchmark—the best we can achieve, but not necessarily applicable in practice. Simulation studies could be conducted to investigate the information loss of the restricted design for blinded trials, which might be negligible especially when the dropout mechanisms of the two groups have similar characteristics. We recognize that, apart from blinding, there may be further reasons for restricting both groups to have the same treatment regimen, eg, when interim analyses are planned or when testing certain underlying assumptions between the two groups. We would argue that, if this was of primary interest, then this should be reflected in the optimality criterion, ultimately leading to a different design. We do note that there is literature that suggests that using a linear mixed model would not be adversely impacted by different designs in the different arms, such as if there were unequal time points, eg, see the work of Hickey et al.<sup>25</sup>

For future research, we suggest considering the impact of an intermittent missing data pattern on the optimal designs for linear mixed models. It may also be interesting to relax the usual convention<sup>5</sup> of automatically selecting the pre-specified end of trial as the latest time point for measurements. In particular, in the situation where dropout is high, an optimal design may select an earlier time point to be the last time point, so that more responses could be observed at this earlier time than at the end of the trial. This could be a feature of an adaptive trial in which the design could be adapted based on interim analysis. In this case, the design framework would need to be extended to include interim updates on the model parameters and hence update the design. Feasibility of this approach would, of course, depend on the research question of the trial.

We note that other missing data analysis approaches, such as multiple imputation or pattern mixture models,<sup>17</sup> could also be applied to the longitudinal data for making inferences. Developing a design framework for these approaches would be substantially more challenging. These problems have not even been tackled yet for the simpler setting of fixed effects models. A further interesting extension of our work would be to consider generalized linear mixed models in the presence of dropouts. The key challenge to considering these suggestions is to find (a good approximation to) the expected information matrix (or to the covariance matrix) for the corresponding models, having accounted for the features of the missing data analysis approach at the design stage of an experiment. Furthermore, more sophisticated optimization algorithms might be required to solve the potentially considerably more complex optimization problems.

## ACKNOWLEDGEMENTS

We would like to thank two unknown reviewers for their careful reading of our work and for useful suggestions that led to substantial improvements of our paper. The first author's research has been funded by the Institute for Life Sciences at the University of Southampton and by the Medical Research Council (grants MR/N028171 and MC\_UP\_1302/4). We would like to acknowledge Clive Holmes, Robert Howard, and Patrick Philips for supplying us with the data from the DOMINO study RCTN49545035, which was funded by the MRC and Alzheimer's Society, UK.

**CONFLICT OF INTEREST**

The authors declare no potential conflict of interests.

**ORCID**

Kim May Lee  <https://orcid.org/0000-0002-0553-973X>

**REFERENCES**

1. Galbraith S, Stat M, Marschner IC. Guidelines for the design of clinical trials with longitudinal outcomes. *Control Clin Trials*. 2002;23(3):257-273.
2. Hedayat A, John PWM. Resistant and susceptible BIB designs. *Ann Stat*. 1974;2(1):148-158.
3. Ghosh S. On robustness of designs against incomplete data. *Sankhyā Indian J Stat Ser B*. 1979;40(3/4):204-208.
4. Tanvir A, Gilmour SG. Robustness of subset response surface designs to missing observations. *J Stat Plan Inference*. 2010;140(1):92-103.
5. Ortega-Azurduey SA, Tan FES, Berger MPF. The effect of dropout on the efficiency of *D*-optimal designs of linear mixed models. *Statist Med*. 2008;27(14):2601-2617.
6. Herzberg AM, Andrews F. Some considerations in the optimal design of experiments in non-optimal situations. *J R Stat Soc Ser B Methodol*. 1976;38(3):284-289.
7. Hackl P. Optimal design for experiments with potentially failing trials. *MODA4 — Advances in Model-Oriented Data Analysis*. Heidelberg, Germany: Physica; 1995.
8. Imhof LA, Song D, Wong WK. Optimal design of experiments with possibly failing trials. *Statistica Sinica*. 2002;12(4):1145-1156.
9. Lee KM, Biedermann S, Mitra R. Optimal design for experiments with possibly incomplete observations. *Statistica Sinica*. 2018;28:1611-1632.
10. Lee KM, Mitra R, Biedermann S. Optimal design when outcome values are not missing at random. *Statistica Sinica*. 2018;28:1821-1838.
11. Ouwens MJNM, Tan PES, Berger MPF. Maximin *D*-optimal designs for longitudinal mixed effects models. *Biometrics*. 2002;58(4):735-741.
12. Berger MPF, Tan FES. Robust designs for linear mixed effects models. *J R Stat Soc Ser C Appl Stat*. 2004;53(4):569-581.
13. Schmelter T. Considerations on group-wise identical designs for linear mixed models. *J Stat Plan Inference*. 2007;137(12):4003-4010.
14. Howard R, McShane R, Lindesay J, et al. Donepezil and memantine for moderate-to-severe Alzheimer's disease. *N Engl J Med*. 2012;366(10):893-903.
15. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592.
16. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc*. 1995;90(431):1112-1121.
17. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. Hoboken, NJ: Wiley-Interscience; 2002.
18. Molenberghs G, Thijs H, Jansen I, et al. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*. 2004;5(3):445-464.
19. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *J R Stat Soc Ser C Appl Stat*. 1994;43(1):49-93.
20. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. *Statist Med*. 2004;23(9):1455-1497.
21. Atkinson A, Donev A, Tobias R. *Optimum Experimental Designs, With SAS*. Oxford, UK: Oxford University Press; 2007.
22. Ibrahim JG, Lipsitz SR, Chen MH. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *J R Stat Soc Ser B Stat Methodol*. 1999;61(1):173-190.
23. Mitra R, Reiter JP. Estimating propensity scores with missing covariate data using general location mixture models. *Statist Med*. 2011;30(6):627-641.
24. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Statistical Science*. 1995;10(3):273-304.
25. Hickey GL, Mokhles MM, Chambers DJ, Kolamunnage-Dona R. Statistical primer: performing repeated-measures analysis. *Interact Cardiovasc Thorac Surg*. 2018;26(4):539-544.

**How to cite this article:** Lee KM, Biedermann S, Mitra R. *D*-optimal designs for multiarm trials with dropouts. *Statistics in Medicine*. 2019;38:2749–2766. <https://doi.org/10.1002/sim.8148>

**APPENDIX****GLOSSARY**

Table A1 presents an explanation for the notations used in this paper. Table A2 gives the details of the class of models considered in this paper. Table A3 explains the restricted and flexible design conditions. Tables A4 and A5 explain the definition of models  $M_g$  and  $M_d$ , respectively.

**TABLE A1** Notation used in this paper

Notation	Definition	Assumption
$\mathbf{y}_i^T = (y_{i1}, y_{i2}, \dots, y_{iq})$	$q$ repeated measurements of subject $i$ .	There are $N$ subjects: $i = 1, \dots, N$
$\delta_k$	Treatment dose of group $k$ .	Could be continuous or categorical. The number of groups, $c$ , is pre-specified.
$\mathbf{t}'_k = \{t_{k1}, t_{k2}, \dots, t_{kq}\}$	A vector of $q$ time points of a subject in group $k$ .	The number of time points, $q$ , is pre-specified. $t_{k1} < t_{k2} < \dots < t_{kq}$
$w_k$	Proportion of subjects in group $k$ .	$0 \leq w_k \leq 1, \sum_{k=1}^c w_k = 1$ .
$\beta$	A vector of fixed effect parameters.	Unknown values that are estimated by the maximum likelihood method.
$\mathbf{X}_i$	(Fixed effect) design matrix of subject $i$ .	Could depend on $\mathbf{t}'_k$ and $\delta_k$ .
$\mathbf{Z}_i$	(Random effect) design matrix of subject $i$ .	Could depend on $\mathbf{t}'_k$ and $\delta_k$ .
$\mathbf{b}_i$	A vector of random effect parameters of subject $i$ .	$\mathbf{b}_i$ follows a multivariate normal distribution with mean zero and covariance matrix $\mathbf{D}$ ; $\mathbf{b}_i, i = 1, \dots, N$ , are independent and identically distributed.
$\mathbf{D}$	Covariance matrix of random coefficients $\mathbf{b}_i$ .	Can be conjectured at design stage, using eg, historical studies.
$\epsilon_i$	A vector of observational errors of repeated measurements on subject $i$ .	$\epsilon_i$ follows a multivariate normal distribution with mean zero and covariance matrix $\sigma^2 \Psi$ ; $\epsilon_i, i = 1, \dots, N$ , are independent and identically distributed, and are independent of $\mathbf{b}_i, i = 1, \dots, N$ .
$\sigma^2 \Psi$	Covariance matrix of observational errors of repeated measurements.	Can be conjectured at design stage, using eg, historical studies.
$\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \sigma^2 \Psi$	Covariance matrix of $\mathbf{y}_i$	
$p_{k,\text{obs}}(t_{kj}, \delta_k)$	Probability of having a response observed on a subject in group $k$ at time point $j$ .	Could depend on $\mathbf{t}'_k$ and $\delta_k$ . Can be conjectured at design stage, using eg, historical studies.
$g_{k,j}$	Number of subjects in group $k$ who have exactly $j$ repeated measurements.	Unknown at design stage.
$m_{k,j} = E[g_{k,j}]$	Expected number of subjects in group $k$ who have exactly $j$ repeated measurements.	Could depend on $\mathbf{t}'_k$ and $\delta_k$ .
$\xi_{D,m}^*$	$D$ -optimal design for a class of model denoted by $m$ .	Maximises the determinant of the information matrix of $m$ .

**TABLE A2** Details of the statistical models considered in this paper

Class of models with random effect parameters, $\mathbf{b}_i = (b_{0i}, b_{1i})^T$ , which have covariance matrix $\mathbf{D} = \begin{pmatrix} \text{var}(b_{0i}) & \text{cov}(b_{0i}, b_{1i}) \\ \text{cov}(b_{0i}, b_{1i}) & \text{var}(b_{1i}) \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$				
Fixed effects model:	Random intercept model:	Random intercept and slope model:	Correlated random intercept and slope model:	
$d_{11} = d_{22} = d_{12} = 0$	$d_{11} > 0, d_{22} = d_{12} = 0$	$d_{11} > 0, d_{22} > 0, d_{12} = 0$	$d_{11}, d_{22} > 0, d_{12} \neq 0$	

**TABLE A3** Explanation of the restricted and flexible design conditions used in the paper

Restricted Design Condition	Flexible Design Condition
$\mathbf{t}'_k$ are the same $\forall k, k = 1, \dots, c$ .	$\mathbf{t}'_k$ could be different.
All subjects are measured at the same time.	Different groups of subjects are measured at different time points.
Suitable for longitudinal studies or blinded trials.	Only suitable for open label or unblinded trials.
Simpler in terms of logistic and administrative arrangements.	Requires different arrangements for different groups.
Less observations might be collected especially when the response rates of the groups are significantly different.	More information could be collected in particularly when time is the major factor that causes dropout.

**TABLE A4** Definition of model  $M_g$ 

Notation	Definition/Usage
$M_g$	$y_i = \mathbf{X}_i \mathbf{K}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i$ , where $\mathbf{K}_i$ is a group indicator matrix. <ul style="list-style-type: none"> <li>For studies that have comparable baseline measurements, and aim to investigate time effect on different groups.</li> <li>Information matrix in the presence of time dependent dropouts is <math>\sum_{k=1}^c \sum_{j=1}^q m_{k,j} \mathbf{K}_k^T \mathbf{X}_{k[j]}^T \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]} \mathbf{K}_k</math>.</li> <li>Design problem is to choose <math>\mathbf{t}'_k</math> and <math>w_k</math>.</li> <li>Section 3.1: Illustration example with <math>c = 2</math>, <math>q = 4</math>. One group has a quadratic <math>p_{1,\text{obs}}(t_{1j}, \delta_1)</math> and the other has a linear <math>p_{2,\text{obs}}(t_{2j}, \delta_2)</math> response probability function. Both response probabilities depend only on the time points. We consider cases with AR(1) serial correlation which have different values of <math>\rho</math>, <math>0 \leq \rho \leq 0.9</math> (and difference of 0.1 between each case) and <math>\sigma^2 = 1</math>. Standardized time design region <math>[-1, 1]</math> that reflects the interval between follow-up visits.</li> <li>Models in illustration have <math>d_{11} = 1</math>, <math>d_{22} = 3</math>, <math>d_{12} = 0.8\sqrt{3}</math> unless the models required some (or all of these) to be set to 0.</li> <li>Figure 1 shows the <math>D</math>-optimal time points with restricted design condition (top row) and flexible design condition (bottom row), respectively.</li> <li>Table 1 shows the maximum and minimum weights across the 10 considered cases of <math>\rho</math>. For <math>\rho &gt; 0</math>, <math>D</math>-optimal designs generally have more experimental units in the group that has a higher response rate within the time region.</li> </ul>

**TABLE A5** Definition of model  $M_d$ 

Notation	Definition/Usage
$M_d$	$y_{ij} = \beta_0 + t_{kj}\beta_1 + \delta_k\beta_2 + b_{0i} + t_{kj}b_{1i} + \epsilon_{ij}$ <ul style="list-style-type: none"> <li>For studies that investigate time effect and treatment effect on the population.</li> <li>Information matrix in the presence of dropouts is <math>\sum_{k=1}^c \sum_{j=1}^q m_{k,j} \mathbf{X}_{k[j]}^T(t, \delta) \mathbf{V}_{k[j]}^{-1} \mathbf{X}_{k[j]}(t, \delta)</math>.</li> <li>Design problem is to choose <math>\delta_k</math>, <math>\mathbf{t}'_k</math> and <math>w_k</math>.</li> <li>Section 3.2: Illustration example with <math>c = 2</math>, <math>q = 5</math>, and a logistic function for response probability that depends on time point and dose level. We consider cases with AR(1) serial correlation, which have different values of <math>\rho</math>, <math>0 \leq \rho \leq 0.9</math> (and difference of 0.1 between each case) and <math>\sigma^2 = 10</math>. The design region for <math>\delta_k</math> is <math>[0, 100]</math> and time region is <math>[0, 365]</math> days.</li> <li>Models in the illustration have <math>d_{11} = 1</math>, <math>d_{22} = 3</math>, <math>d_{12} = 0.8\sqrt{3}</math> unless the models required some (or all of these) to be set to 0. Figure 2 shows the <math>D</math>-optimal time points (first row), optimal dose (second row), and optimal weight (third row), with restricted design condition.</li> <li>Section 4 shows how to verify some potential optimal designs for a case study. Table 3 presents the <math>D</math>-optimal time points and weights for various version of this model, which have five time points and four time points, respectively. Table 4 shows the simulation performance of the design candidates.</li> <li>Models in Section 4 consider <math>\{\hat{\rho}, \hat{d}_{11}, \hat{d}_{22}, \hat{d}_{12}\} = \{0.3326, 2.661^2, 0, 0\}</math>, <math>\{0.3326 \times 2, (2.661 \times 2)^2, 0, 0\}</math>, <math>\{0.3326, 2.661^2, 2, 0\}</math>, and <math>\{0.3326, 2.661^2, 2, -1\}</math>, respectively.</li> </ul>